Available online at www.jeeir.com

JEEIR

e-ISSN: 2289-2559

Journal of Emerging Economies & Islamic Research 6(1) 2018, 1 – 5.

Journal of Emerging Economies and Islamic Research

Viewpoint by Guest Writer

Managing economic and Islamic research in big data environment: From computer science perspective

Nordin Abu Bakar*

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Malaysia

ARTICLE INFO

Article history:
Received 12 September 2017
Received in revised form
14 September 2017
Accepted 17 September 2017
Published 30 September 2017

Keywords:
Big data
Machine learning
Algorithm

ABOUT THE GUEST WRITER

Nordin Abu Bakar is an associate professor of Computer Science from the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam. He earned his PhD in Computer Science from the University of Essex, England in 2001. His research interest and expertise include software engineering, evolutionary computation, genetic algorithms, islamic finance, computer science education and machine learning.

ABSTRACT

Research in economic and Islamic fields are facing a major challenge in the surge of big data. The landscape and the environment produce problems of massive magnitude and demand robust solutions. The traditional method might not be able to cater for this huge challenge; so, researchers must embark on the mission to seek new and versatile methods to solve the complex problem. If not, the research output would end up with sub-optimal results. In computer science, there are machine learning algorithms that have been used to solve problems in a such complex environment. This article explains the current demanding situation facing many researchers and how those algorithms have successfully solved some of the problems. The potential applications of the methods should be learned and utilised to

* Corresponding author. Tel.: +603-5521-1162
E-mail address: nordin@fskm.uitm.edu.my

1. Introduction

A major change has occurred in many research fields as the phenomenon of big data overwhelms the research landscape. The economic and Islamic research has no exception and has been affected in a big way. The data have become huge and complex in terms of volume, velocity, variety and veracity. These new characteristics of data that the researchers have to deal with requires them to be equipped with robust methods and tools. In computer science, machine learning methods and tools have facilitated research since the early days of computing. The methods have gone through refinements and tests to ensure their robustness remain relevant in this big data environment. The amount of information that researchers have to handle has been increased beyond the capacity and capability of the current research tools. Data from social media and smart phones in terms of video and audio formats have exploded into the existing database. This phenomenon has created new insights into the dimensions of data analytics. The opportunity to understand the data through this new platform must be harnessed properly by using appropriate data analytic tools. It would improve analysis, assumptions and predictions as well as driving innovations in the economics and Islamic research. Researchers have to face with uncleaned and unstructured data far too often, and now the big data phenomenon has added a new challenge for them to manage. The solution, I believe, lies in the area of machine learning in Computer Science.

This article describes the methods and tools available in machine learning that will help to facilitate economic and Islamic research in big data environment. Understanding halal and haram

2. Big data

This term refers to the huge information available on the internet that can be used to understand issues in emerging economics and Islamic research. As mentioned above, the big data phenomenon exploded due to four characteristics: volume, variety, velocity and veracity.

- ***Volume*** refers to the fact that enterprises of all types are awash with ever-growing data of all types, easily amassing terabytes - even petabytes - of information. Trying to analyse data of this scale is a challenge in itself.
- ***Variety*** signifies that big data could be any type of data - structured and unstructured data such as text, sensor data, audio, video, click streams, log files and more. New insights can be found when analysing these data types, particularly when different data sets are combined.
- ***Velocity*** is concerned with the time value of data; sometimes even ten seconds is too late. For time-sensitive processes such as catching fraud, this data must be used as it streams into your enterprise in order to maximise its value. A telecommunications organisation, for example, might analyse 50 million call detail records per day in real-time to predict customer churn faster. Or a credit card company might scrutinise five million trading events per day to identify potential fraud.
- ***Veracity*** refers to the truthfulness of data. As the speed of data changing and spreading very fast, the truthfulness of the information might evolve along with the sentiment that they carry.

Examples

Machine learning has been used and proven to perform effectively in many business applications such as web search[9], advertisement placement[10], credit scoring [9], data analysis[11], financial analysis[11], market prediction [12], gene analysis [12], behavior analysis [13], weather forecasting [14], and big data analytics [8], [5], [15]. In management, machine learning algorithms are designed to learn important attributes to use for making the decisions [16]. In banking sectors and financial technology, for example, machine learning helps the banking sectors to make decision whether to provide the financial loan or not to a company or individual with some degrees of risks [17]. Bankruptcy prediction helps financial institution to assess the individual or company finance data, examine the probability of

bankruptcy and make decision to give the loan or not. False decision will put the institution at risk [18], [20]. Recently, many studies in machine learning proved that machine learning technique has been successfully applied in bankruptcy prediction with high accuracy [6]. Research on using machine learning to determine bankruptcy prediction has been dated since the 1990's.

The table below shows an adaptation of previous works done on bankruptcy prediction since the 1900's based on [5].

Table 1. Methods used on bankruptcy prediction

Period	Disciminant Analysis	Logit analysis	Probit analysis	Neural network	Others
1960's	2	0	0	0	1
1970's	22	1	1	0	4
1980's	28	16	3	1	7
1990's	9	16	3	35	11
2000's	2	3	0	4	3
Overall	63	36	7	40	26

As stated in the table above, neural networks gained popularity starting from the early 1990's until today for prediction of bankruptcy. Examples of articles that investigate the use of artificial intelligence include research articles [1][2][3][4] and [5] where the machine learning techniques involved were artificial neural networks, back propagation networks and multilayer perceptron. Hence, the study of machine learning for bankruptcy prediction is not something new itself. Based on the article of [5] neural networks can predict bankruptcy with the highest accuracy of more than 90% and the lowest accuracy was at 71% in which the author suggested that MDA and neural networks can be the most promising methods to act as a bankruptcy prediction model.

Machine learning methods and models are described in terms of algorithms. They use a variety of techniques to solve a problem from different domain. Some examples of popular machine learning algorithms are introduced in Table 2 below.

Table 2. Examples of machine learning algorithms

Algorithms	Description
C4.5	A decision tree method that trains data for classification and regression problems. Tree structures are constructed based on the actual values of attributes in the data and develop until a prediction decision is made for a given record.
K-Means	This clustering method uses the inherent structures in the data to best organize the data into groups of common features.
SVM	Support Vector Machine (SVM) uses classification and regression analysis to analyse data.
Apriori	Uses association to extract the best rules that explain the observed relationships between variables in the data set.
EM	Expectation Maximization (EM) is a clustering method that describes the class of problem and the class of methods.
PageRank	This is a search ranking algorithm using hyperlinks on the web. The search engine, Google, was developed based on this algorithm.
AdaBoost	It is one of the ensemble algorithms which is composed of multiple weaker models that are independently trained and whose predictions are combined in some way to make the overall prediction. Other similar models are random Forest and boosting.

Algorithms	Description
kNN	k-Nearest Neighbour (kNN) falls under instance based learning model. Using a database, kNN compares new data with it and produces similarity measure to find the best match and make a prediction. Other algorithms in this category include Self-Organizing Maps(SOM) and Learning Vector Quantization(LVQ).
Naïve Bayesian	Uses Bayes' Theorem for classification and regression problems.
CART	Classification and Regression tree (CART) works with decision tree structure to arrive at the prediction decision.

3. Conclusion

The research landscape has changed due to the explosive nature of data dynamics from the internet. The economic and Islamic research have no exception and must be scaled to the environment. The complexity of the environment demands versatile and robust tools to analyse the data and produce meaningful results. There are many tools available for the researchers to choose from; depending on the type of data being analysed. Traditional statistical tools are the most common but some methods from the area of computer science have started to contribute for some interesting results and better performance. These machine learning methods that are used in computer science research have also been tested on finance and business data. The results have shown positive outcomes and proved to be useful to overcome the complexity of the big data environment.

References

[1] A. Mukhopadhyay, S. Tiwari, A. Narsaria and B. Karmaker, "A New Approach to Predicting Bankruptcy: Combining DEA and Multi-Layer Perceptron", International Journal of Computer Science, vol. 9, no. 2, pp. 1-6, 2012.

[2] E. Alfaro, N. Garcia, M. Gamez and D. Elizondo, "Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks", Decision Support Systems, vol. 45, no. 1, pp. 111-121, 2007.

[3] M. Odom and R. Sharda, "A Neural Network Model for Bankruptcy Prediction", 1990, pp. 163-167.

[4] G. Zhang, M. Y. Hu, B. Eddy Patuwo and D. C. Indro, "Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis", European Journal of Operational Research, vol. 116, no. 1, pp. 16-32, 1999.

[5] S. Huang, C. Tsai, D. Yen and Y. Cheng, "A hybrid financial analysis model for business failure predicition", Expert systems with applications, vol. 35, no. 3, pp. 1034-1040, 2008.

[6] X. Ding, Y. Zhang, T. Liu and J. Duan, "Deep Learning for Event-Driven Stock Prediction", 2016, pp. 2327-2332.

[7] J.A Sirignano, A. Sadhwani and K.Giesecke, " Deep Learning for Mortgage Risk", 2016.

[8] Q. He, N. Li, W. J. Luo, and Z. Z. Shi, ―A survey of machine learning algorithms for big data,‖ Moshi Shibie yu Rengong Zhineng/Pattern Recognit. Artif. Intell., vol. 27, no. 4, pp. 327–336, 2014. [5] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, ―Deep learning applications and challenges in big data analytics,‖ J. Big Data, vol. 2, no. 1, p. 1, 2015.

[9] J. Boyan, D. Freitag, and T. Joachims, ―A Machine Learning Architecture for Optimizing Web Search Engines,‖ Aaai, no. 3, pp. 1–8, 1996.

[10] D. Hillard, S. Schroedl, E. Manavoglu, H. Raghavan, and C. Leggetter, ―Improving ad relevance in sponsored search,‖ Proc. third ACM Int. Conf. Web search data Min. - WSDM '10, p. 361, 2010. [9] S. Lessmann, B. Baesens, H. V. Seow, and L. C. Thomas, ―Benchmarking state-ofthe-art classification algorithms for credit scoring: An update of research,‖ Eur. J. Oper. Res., vol. 247, no. 1, pp. 124–136, 2015. 11

[11] R. Burbidge, M. Trotter, B. Buxton, and S. Holden, ―Drug design by machine learning: support vector machines for pharmaceutical data analysis,‖ Comput. Chem., vol. 26, no. 1, pp. 5–14, 2001. [11] J. Perols, ―Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms,‖ Audit. A J. Pract. Theory, vol. 30, no. 2, pp. 19–50, May 2011.

[12] W. S. Hayes and M. Borodovsky, ―How to interpret an anonymous bacterial genome: machine learning approach to gene identification.,‖ Genome Res., vol. 8, no. 11, pp. 1154–71, Nov. 1998.

[13] S. Saad et al., ―Detecting P2P botnets through network behavior analysis and machine learning,‖ in 2011 Ninth Annual International Conference on Privacy, Security and Trust, 2011, pp. 174–180.

[14] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, ―Predicting solar generation from weather forecasts using machine learning,‖ in 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), 2011, pp. 528–533.

[15] T. Condie, P. Mineiro, N. Polyzotis, and M. Weimer, ―Machine learning for big data,‖ Proc. 2013 Int. Conf. Manag. data - SIGMOD '13, p. 939, 2013.

[16] I. H. Witten, E. Frank, and M. a Hall, Data Mining: Practical Machine Learning Tools and Techniques (Google eBook). 2011.

[17] S. Karlos, S. Kotsiantis, N. Fazakis, and K. Sgarbas, ―Effectiveness of semi-supervised learning in bankruptcy prediction,‖ 2015.

[18] F. Babič, C. Havrilová, and J. Paralič, ―Knowledge Discovery Methods for Bankruptcy Prediction,‖ BIS, pp. 151–162, 2013.

[19] C.-H. Chou, S.-C. Hsieh, and C.-J. Qiu, ―Hybrid genetic algorithm and fuzzy clustering for bankruptcy prediction,‖ Appl. Soft Comput., vol. 56, pp. 298–316, 2017.

[20] A. Mukhopadhyay, S. Tiwari, A. Narsaria, and B. R. Karmaker, ―A New Approach to Predicting Bankruptcy : Combining DEA and Multi - Layer Perceptron,‖ Int. J. Comput. Sci. Issues, vol. 9, no. 4, pp. 71–78, 2012